ICIC Express Letters
Volume **2**, Number **1**, March **2008**
 ISSN 1881-803X
pp. 65–72

VIDEO TRACKING SYSTEM: A SURVEY

Zalili Musa and Junzo Watada

Graduate School of Information, Production and System
Waseda University
2-7 Hibikino, Wakamatsu, Kitakyushu 808-0135, Japan
zalili@ump.edu.my; junzow@osb.att.ne.jp

Received December 2007; accepted February 2008

Abstract. *A video tracking system raises a wide possibility in today's society. This system is used in various applications such as security, monitoring, robotic, and nowadays in day-to-day applications. However the video tracking systems still have many open problems and many of research activities in a video tracking system are explored. Therefore in this article, an overview of recent research activities in a video tracking systems has been presented. In our survey, we have discussed the important issues related to a video tracking system including current issues, methodology and future view of tracking system. In methodology section, we provide detailed description of four main tasks in tracking system: detection, tracking, object representation and recognize object behavior.*
Keywords: Video tracking system, Detection, Tracking, Object representation, Object behavior, Recognition

1. **Introduction.** A video tracking system raises a wide possibility in today's society. This system is applied to such security systems as real-time temperature scanning during a SARS eruption in East Asia; a surveillance system that is used in a shopping mall, ATM, booth railway station and airport; a monitoring system, for example, traffic monitoring in a highway and nowadays has been used widely on day-to-day application such as a video camera to a cell phone. Based on Rita Cucchiara, the projection of the compound annual growth rate of video surveillance markets is about 23% over 2001-2011, to touch US$670.7 million and US$188.3 million in USA and Europe, respectively [1]. According to the increase of its use, this means rapidly such technologies invaded in the society.

Therefore, many of research activities in a video tracking system are explored. In early 1980s, image processing started to be employed in security technology when a system of Video Motion Detection (VMD) was introduced. Since then, many researchers have developed various methods, algorithms and techniques that can be implemented in the video tracking system such as [2-5]. However, the performance of tracking algorithms using image processing method depends on how well they can estimate the visual information such as an articulate object structure and due to the change of a view point and illumination [6].

From the above analysis, the objective of this paper is to provide the present state of arts and discuss a future trend to research on a tracking system. In this paper, three main fields of tracking systems are outlined. In Section 2, we describe briefly the overview of video tracking system. In Section 3, we discuss the methodology of tracking systems, where we provide detailed description of four main tasks in tracking system: detection, tracking, object representation and recognize an object behavior. Section 4 presents the future view of tracking systems and finally, conclusion is given in Section 5.

2. **Video Tracking Systems.** Video tracking can be defined an action to estimate the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video.

At the early stage of tracking systems, a lot of researchers focused on tracking an object movement in an indoor environment such as Q. Cai and J. K. Aggarwal, they have developed tracking human motions using a plural fixed camera. They have focused on an indoor environment and multivariate Gaussian models are applied to find the most likely matches of human subjects [7]. Furthermore, Sarma and Jakup developed a human tracking system in two approaches of a method. The first approach is to compare the computed motion trajectories to the reference trajectories. Meanwhile, in the second approach they identified specific discrete events such as line crossings and compared sequences of these evens to reference sequences. But in their research, they tested their proposed method in an indoor environment [8].

Basically, tracking a human motion in an indoor environment is easier than in an outdoor environment. This is because, the indoor environment has other features such as polygons, lines or other line structures in the background that can be used for simplifying the process of recognition [9]. Besides that, the curved contour or shape of a human can reduce a possibility of mismatching the object with the background [8].

However, because of demanding from a society, tracking systems are widely used for outdoor environment activities. Based on Chia Jung pai et al., their research is focusing on pedestrian detection and tracking at crossroads. Their research has used a pedestrian model and a walking rhythms approach to recognize a single pedestrian [10]. Meanwhile, Zhijun Qui et al., have proposed corner feature extraction, motion matching and object classification for the detection of a pedestrian and bicycle on a road [11]. Jose Melo et al. have studied detection and classification of highway lanes using vehicle motion trajectories [12]. That research is used an uncelebrated stationary or PTZ cameras to analyze traffic behaviors in real time.

This situation is so different from tracking object motions in the indoor environment. In an outdoor environment, tracking an object motion is a problem how to deal with a background itself. The background of a moving picture should be different regarding to situations and places even though data are taken in the same period. Besides that, the color, pattern or shape of a background sometimes resemble with the shape, color or pattern of an object that we need to track.

On the other hand, the tracking system still have many open problems should be solved such as depth information for tracking. In order to detect an unsafe situation, it is required to recognize a natural situation, environment or human and vehicle behaviors in a whole time-length and in a spatial universe such as provide a views taken by cameras scattered and placed in various positions or from various angles in buildings, campus, airports, stations, streets, highways, or towns for tracing environments, vehicles and humans.

Therefore, to solve that problem, nowadays most of new research activities in a tracking system are exploring larger dimension for tracking systems [1]. As we know, many cameras are employed in a distributed video surveillance system to view, watch and recognize irregular or unsafe behaviors. For example, it is not easy to track a human behavior over many cameras, each of which consists a limited partial space in a vast universe [13]. Therefore, a multi-camera approach can be an alternative to solve this problem.

Using a multi-camera tracking system the real-time processing is one of the most important issues to be considered. As we know, using multiple cameras the system will record a huge volume of video data and for each video have a different angle of view. Therefore a challenging task is to track an object movement from a multi video stream. Regarding on the issue that it is reliable to ensure the application can track an object movement on multi video data quickly under minimizing processing time.

Generally, in video tracking systems, visual content extraction and image understanding are a crucial issue and still partially unsolved for tracking systems such as enlarge review, enhance review, multiple review and so on. How to overcome the above problem, in the next section we will discuss the detail of methodology.

3. **Methodology of Tracking System.** Currently, researchers have proposed various methodology of tracking applications. Alper Yilmaz et al. concluded on their survey that three components are important tasks in tracking systems: detection, tracking and object representation [14]. However, some of researchers identified that besides these three components, another important component should be considered as to recognize understand an object behavior or object activity.

3.1. **Detection.** Based on David Beymer and Kurt Konolige, the challenge in a tracking system is to find an efficient method for detecting and tracking an object under fairly difficult natural conditions. In the tracking system, detection means process of localization of object movement or allowing the system to extract and log the person's (object) appearance in the repository. Normally, detection process is required during detecting the first object enters in the frame or needed to detect an object in every frame [15]. These two situations are depending on the requirement of applications.

Currently, various detection methods have been proposed for tracking systems. Based on Robert T. Collins et al. the moving target detection can be divided by three conventional approaches: temporal differencing, background subtraction and optical flow [16]. Meanwhile based on Alper Yilmaz et al. it is popular that four categories of a detection method are used in tracking systems which are point detectors, segmentation, background subtraction and supervised learning [14]. Generally, background subtraction is mostly used in tracking systems. Even though, this method is extremely sensitive to dynamic scene changes for example lighting or inappropriate event.

Regarding our study many of researchers have their own methods to solve the problem of object detection. For examples David Beymer and Kurt Konilige used disparity templates for the detection process of humans [15]. Robert T. Collins et al. used dynamic background subtraction method to detect the movement of objects at urban area [16].

Chia Jung Pai et al. are focusing on pedestrian detection and tracking at crossroads. Their research has used a pedestrian model and walking rhythms to recognize a single pedestrian [10]. Meanwhile, Zhijun Qui et al. have proposed corner feature extraction, motion matching and object classification for the detection of a pedestrian and bicycle on a road [11].

Meanwhile, in robust real-time upper body limb detection, they used color skin and edge information as a visual feature to detect the limbs of a human [17]. Musa et al. used a shadow based method to recognize the location of vehicles in a highway. In other researches on the multi-camera tracking system for human motions in an outdoor environment, they used histogram method to identify the location of humans [18]. Nowadays, hybrid methods have been used to solve the problem of object detection.

3.2. **Tracking.** The objective of tracking method is to generate the path of an object by locating an object position in every frame from a video stream. Basically, two main purposes of tracking method are to determine when a new object enters the system and secondly, to estimate the position of an object over time.

Generally tracking object movements is a challenging task because a target changes dynamically. In a traditional way, an object motion can be measured by two models: stochastic models with deterministic and random components and secondly stochastic model by classical deterministic mechanics [19]. Figure 1 shows as examples of such that a model includes constant position, constant velocity and constant acceleration. In tracking systems, to track an object motion based on constant velocity and constant acceleration model it is difficult because this model is not a linear motion and also will affect the size of an object during movement. Therefore we need a compatible method to overcome this problem.

Various techniques have been developed for an object tracking system. A. Yilmaz et al. have classified object tracking methods into three categories: firstly point tracking,

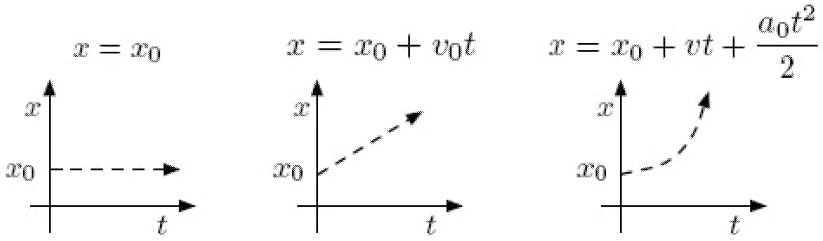

FIGURE 1. Traditional motion models [19]

secondly kernel tracking and lastly silhouette tracking [14]. Nowadays many of researchers have proposed their own method. For example, in real-time tracking of plural people using continuous detection, David Beymer and Kurt Konilige used Kalman filtering technique to track people's movement [15]. However, based on Isard and Blake, this technique has limited use because they are based on unimodal Gaussian densities that cannot support simultaneous alternative motion hypotheses [20].

Hwasup Lim et al. have studied dynamic appearance modeling for the human tracking system. This method used identification of the robust system dynamics and nonlinear dimensionality reduction technique to track a human [21].

In robust real-time upper body limb detection and tracking, Matheen Siddiqui and Gerard Medioni have designed a real time and robust system to detect or track the limbs of a person and then track the forearms of the found limbs [17]. In the research, information of color's skin and edge is used as a visual feature to detect the limbs of a human. After the detection process, in order to track the limbs in the next frame they used smoothness assumption that the limb in the next frame must be near to the limb found in the previous frame.

Meanwhile Robert T. Collins et al. used a simple approach based on a frame-to-frame matching cost function to track the object movement such as humans and vehicles [16]. Chun Che Fung and Nicholas Jerrat proposed a neural network based intelligent intruder's detection and tracking system using CCTV images used barycentre calculation technique in a tracking module. This technique can be loosely compared to the center of gravity of an object [22].

3.3. **Object representation.** Object representation is one of the most important tasks in tracking system. In a tracking scenario an object can be represented as shape, point and forth depending on the application. However, the object representation can be categorized using such features as point, primitive geometric shape, object silhouette and counter, articulated shape models and skeletal models [14]. We can use this category as a shape representation.

Meanwhile an object tracking also can be represented as object appearances such as probability densities of object appearance, template, active appearance models and multiview appearance models [14]. An object appearance means some features or attributes of an object are considered to represent an object such as color, texture and so on. Furthermore, an object appearance also can be called as appearance representations or appearance features.

In general, many of researchers used a shape representation to express an object movement. For example, Ismail Haritaoglu et al. used silhouette, rectangular and centroid appearance to represent a human movement [23]. Yasuyuki Matsushita et al. used a shape rectangular to represent an object in their research [24]. In detection and classification of highway lanes using vehicle Motion Trajectories, they used rectangular shape and probability densities of an object appearance to represent vehicles motion [12].

C. Micheloni and G. L. Foresti used a multiple point to represent the interest points of objects [25]. Meanwhile, Qiang Wang et al. used control points on object contour such as fish shape or lips shape appearance in their research [26]. Michael Isard and Andrew Blake also used control points on object contour for representing a head motion [20].

3.4. **Recognize an object behavior.** As we mentioned above, detection, tracking and object representation are a main module in tracking system and a part of module in surveillance and monitoring system. But these modules are not sufficient in obtaining the final result to the end user. Therefore, we need a system that can recognize an object behavior to ensure the system can give the decision making

Recognize an object behavior or object activity has been addressed by many researchers in different fields such as computer vision, multimedia processing and pervasive computing. Behaviors refer to the actions or reactions of an object or organism, usually in relation to the environment. Meanwhile human behavior is the collection of activities performed by human beings. Therefore, to understand an object behavior is one of most difficult open problem in tracking application. After detection and tracking an object, a tracking system should determine what kind of activity the object is doing. For example used star skeletonization procedure to analyze the human motion. Using this procedure, they can distinguish a human motion between walking and running [16]. Based on Sangho Park et al. used a hierarchical Bayesian network (BN) to recognize of two person interactions. Where using a low level of BN can track a part of a human body and to track a whole body, they used a high level of the BN [27].

In statistical synthesis of facial expression, Lisa Gralewski et al. used principal components analysis (PCA) and the application of an auto-regressive process (ARP) to define a set of emotion model. This technique generates a video texture from sequences of coherent facial expression and head motions [28]. Meanwhile, Hatice Gunes et al. interpreted a human movement, body behavior and facial expression in order to make a human-computer interface truly nature [29].

Marco Leo et al, used Hidden Markov Models (HMM) to recognize of human behavior. In this research, they can recognize the human's postures such as a human standing, squatted and bent. Using Hidden Markov Models can identify 4 kinds of activities: 1. walking, 2. Probing the subsoil by a stick, 3. Damping the ground with a tank and 4. Picking-up some objects from the ground [30].

Basically for vehicle tracking system, many of researchers recognize a vehicle behavior based on color, shape, vehicle's plate and vehicle's motion such as turn right and left or depend on a road. For example Musa et al, using a car plate to track a vehicles movement on the highway [13]. In real-time video surveillance for traffic monitoring using virtual line analysis, Belle L. Tseng et al. used type and color of vehicles to clarify them [31].

4. **Future Research.** A video tracking system has a wide possibility in today's society. This system is dealt in biometrics [32] and situation/context awareness technologies [33]. Today biometric technology is widely used in a real world. The production of biometric apparatuses increases exponentially from 1996 [33] as shown in Figure 2. Contrarily, the price of each apparatus decreases linearly each year. According to the increase of its use, the price in 2000 is one tenth less than one in 1990. This means rapidly such technologies invaded in our society.

But such researches have very partly and individually been pursued. In order to detect an unsafe situation, it is required to recognize a natural situation, environment or human and vehicle behaviors in a whole time-length and in a spatial universe. One method is to provide a tracking system employing views taken by cameras scattered and placed in various positions or from various angles in buildings, campuses, airports, stations, streets, highways, or towns for tracing environments, vehicles and humans.

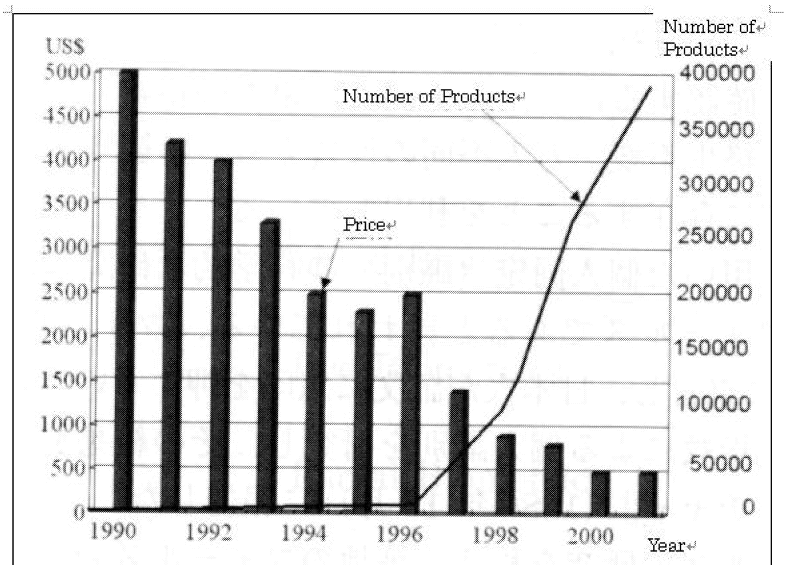

FIGURE 2. Price and product amount of biometric apparatus [33]

Therefore to overcome crucial issues, nowadays most of new research activities in a tracking system are exploring larger dimension for tracking systems [1]. However, in a real-time for video surveillance, visual content extraction and image understanding are a crucial issue and still partially unsolved for video streams Furthermore, the creation, processing and management of these data types require an enormous computational effort, often too high for a single processor architecture. Therefore, we need a parallel and distributed computing in order to speed up the process of image and video application as a solution.

5. **Conclusions.** In this paper, an overview of recent research activities in a video tracking systems has been presented. In video tracking systems, the researchers will cope with the problems of visual content extraction and image understanding that several of techniques have been proposed in detection, tracking, object representation and recognize human behaviors. Furthermore, for the future of video tracking systems, enlarge and enhance the view of a tracking system with multiple cameras should be considered. In this issue we should deal with environment situation in different object views, a huge volume of video data recorder and real-time processing.

Acknowledgment. The first author would like to thank to Universiti Malaysia Pahang (UMP) and Kementerian Pengajian Tinggi Malaysia for a study leave.

REFERENCES

[1] R. Cucchiara, Multimedia surveillance systems, *Proc. of the Third ACM International Workshop on Video Surveillance & Sensor Networks*, Hilton, Singapore, pp.3-10, 2005.

[2] A. D. Bagdanov, A. del Bimbo and F. Pernici, Acquisition of high-resolution images through on-line saccade sequence planning, *Proc. of the Third ACM International Workshop on Video Surveillance & Sensor Networks*, Hilton, Singapore, pp.121-130, 2005.

[3] S. Uchihashi, Improvising camera control for capturing meeting activities using a floor plan, *Proc. of the Ninth ACM International Conference on Multimedia*, Ottawa, Canada, pp.12-18, 2001.

[4] Y.-F. Wang, E. Y. Chang and K. P. Cheng, A video analysis framework for soft biometry security surveillance, *Proc. of the Third ACM International Workshop on Video Surveillance & Sensor Networks*, Hilton, Singapore, pp.71-78, 2005.

[5] Z. Naor, *Tracking Mobile Users with Uncertain Parameters, Wireless Networks 9*, Kluwer Academic Publishers, Netherlands, 2003.

[6] A. Utsumi, H. Yang and J. Ohya, Adaptive human motion tracking using non-synchronous multiple viewpoint observations, *Proc. of the International Conference on Pattern Recognition*, vol.4, pp.607-610, 2000.

[7] Q. Cai and J. K. Aggarwal, Tracking human motion using multiple cameras, *Proc. of the International Conference on Pattern Recognition*, vol.3, pp.68-72, 1996.

[8] S. Pingali and J. Segen, Performance evaluation of people tracking systems, *Proc. of the Workshop on Applications of Computer Vision*, pp.33-38, 1996.

[9] Q. Cai, A. Mitiche and J. K. Aggarwal, Tracking human motion in an indoor environment, *Proc. of the International Conference on Image Processing*, vol.3, pp.68-72, 1995.

[10] C.-J. Pai, H.-R. Tyan, Y.-M. Liang, H.-Y. M. Liao and S.-W. Chen, Pedestrian detection and tracking at crossroads, *Proc. of the International Conference on Image Processing*, pp.II-101-4, vol.3, 2003.

[11] Z. Qui, D. Yao, Y. Zhang, D. Ma and X. Liu, The study of the detection of pedestrian and bicycle using image processing, *Intelligent Transportation Systems*, vol.1, pp.340-345, 2003.

[12] J. Melo, A. Naftel, A. Bernardino and J. Santos-Victor, Detection and classification of highway lanes using vehicle motion trajectories, *IEEE Transactions on Intelligent Transportation Systems*, vol.7, pp.188-200, 2006.

[13] Z. B. Musa and J. Watada, A grid-computing based multi-camera tracking system for vehicle plate recognition, *Kybernetika Journal*, Czech Republic, vol.42, no.4, pp.495-514, 2006.

[14] A. Yilmaz, O. Javed and M. Shah, Object tracking: A survey , *ACM Computing Surveys*, vol.38, no.13, pp.1-45, 2006.

[15] D. Beymer and Konolige, Real-time tracking of multiple people using continuous detection, *Proc. of the International Conference on Computer Vision*, 1999.

[16] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto and Hasegawa, System for video surveillance and monitoring, *Technical Report CMU-RI-TR-00-12*, Robotics Institute, Carnegie Mellon University, 2000.

[17] M. Siddiqui and G. Medioni, Real time limb tracking with adaptive model selection, *Proc. of the International Conference on Pattern Recognition*, vol.4, pp.770-773, 2006.

[18] Z. B. Musa and J. Watada, Dynamic tracking system for object recognition, *Proc. of the International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, vol.4694/2007, pp.378-385, 2007.

[19] G. Welch, B. D. Allen, A. Ilie and G. Bishop, Measurement sample time optimization for human motion tracking/capture systems, *Proc. of the Trends and Issues in Tracking for Virtual Environments, Workshop at the IEEE Virtual Reality 2007 Conference*, Hilton Charlotte Center City, Charlotte, North Carolina, USA, 2007.

[20] M. Isard and A. Blake, Contour tracking by stochastic propagation of conditional density, *Proc. of the European Conference on Computer Vision*, pp.343-356, 1996.

[21] H. Lim, O. I. Camps, M. Sznaier and V. I. Morariu, Dynamic appearance modeling for human tracking, *Proc. of the Conference on Computer Vision and Pattern Recognition*, vol.1, pp.751-757, 2006.

[22] C. C. Fung and N. Jerrat, A neural network based intelligent intruders detection and tracking system using CCTV images, *Proc. of the TENCON*, vol.2, pp.409-414, 2000.

[23] I. Haritaoglu, D. Harwood and L. S. Davis, W4: Real-time surveillance of people and their activities, *Pattern Analysis and Machine Intelligence*, pp.809-830, 2000.

[24] Y. Matsushita, K. Nishino, K. Ikeuchi and M. Sakauchi, Shadow elimination for robust video surveillance, *Proc. of the Workshop on Motion and Video Computing*, pp.15-21, 2002.

[25] C. Micheloni and G. L. Foresti, Fast good features selection for wide area monitoring, *Proc. of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pp.271-276, 2003.

[26] Q. Wang, G. Xu and H. Ai, Learning object intrinsic structure for robust visual tracking, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, vol.2, pp.II-227-II-233, 2003.

[27] S. Park and J. K. Aggarwal, Recognition of two-person interactions using a hierarchical Bayesian network, *Proc. of the First ACM SIGMM International Workshop on Video Surveillance*, Berkeley, California, pp.65-76, 2003.

[28] L. Gralewski, N. Campbell, B. Thomas, C. Dalton and D. Gibson, Statistical synthesis of facial expressions for the portrayal of emotion, *Proc. of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, Singapore, pp.190-198, 2004.

[29] H. Gunes, M. Piccardi and T. Jan, Face and body gesture recognition for a vision-based multimodal analyzer, *Proc. of the Pan-Sydney Area Workshop on Visual Information Processing*, pp.19-28, 2004.

[30] M. Leo, T. D'Orazio and P. Spagnolo, Human activity recognition for automatic visual surveillance of wide areas, *Proc. of the ACM 2nd International Workshop on Video Surveillance and Sensor Networks*, New York, USA, pp.124-130, 2004.

[31] B. L. Tseng, C.-Y. Lin and J. R. Smith, Real-time video surveillance for traffic monitoring using virtual line analysis, *Proc. of the 2002 IEEE International Conference on Multimedia and Expo*, vol.2, pp.541-544, 2002.

[32] Y. Seto, Trend of biometric security technology, *J. Soc. Instrum. Control Engineering*, Special Issue: Advances in Biometric Identification, vol.43, no.7, 2004.

[33] T. Sato, Technical view: Situation recognition and its future in ubiquitous society – Hhuman support systems in terms of environmental system and contents system –, *J. Systems Control Inform.*, vol.49, no.4, 2005.

[34] Y. Sawada, Risk-sensitive tracking control of stochastic systems with preview action, *Int. J. of Innovative Computing, Information and Control*, vol.4, no.1, pp.189-198, 2008.

[35] A. Fekih, H. Xu and F. Chowdhury, Neural networks based system identification techniques for model based fault detection of nonlinear systems, *Int. J. Innovative Computing, Information and Control*, vol.3, no.5, pp.1073-1085, 2007.